PDR RID Report

Date Last Modified 6/20/95
Originator Bill Emery/Gail McConaughy **Phone No**
Organization
E Mail Address
Document
Section **Page** **Figure** **Table**

RID ID **PDR** 159
Review CSMS
Originator Ref
Priority 1

Category Name System-level **Actionee** HAIS
Sub Category Subsetting vs Network Cost
Subject Remote file mounting - Distributed File Access

Description of Problem or Suggestion:

Analyze impacts of science user remote mounting of data files (potential future user requirement):

1. What are science user communities currently doing in this area, and what would they like to do?
2. What are potential impacts on SDPS and CSMS of this paradigm (e.g., network bandwidth)? What should ECS plan for?
3. What prototyping is needed and planned for SDPS and CSMS in EPs and end-to-end testbed?
4. What trade studies are neeeded and planned?

Would a distributed file access method for delivery of data between DAACs impact bandwidth sizes (e.g. direct file access by a remote ingest process)?

Originator's Recommendation

1) Examine the feasibility and impact of a distributed file access approach. Determine impacts if feasible.
2) Define a validation process and execute it.
CDR sizing of DAAC-to-DAAC links should be changed based on the above actions as determined by the analysis.

GSFC Response by: **GSFC Response Date**

HAIS Response by: Forman **HAIS Schedule** 2/24/95
HAIS R. E. Armstrong **HAIS Response Date** 6/5/95

The following are responses to each item of the RID:

1. As part of its work with the science users to understand the needs of the scientific community, the ECS User Characterization Team has developed, through site visits and interviews with the scientists, operational scenarios to depict user interaction with ECS and the high-level functionality to support such needs. In fact, four of the 27 science user scenarios, use functionality that may require remote file access. The four scenarios were originated by Bruce Barkstrom (#13), Walter Rosenthal (#14), Raul Lopez (#15), and Leonard Walstad (#16). These scenarios provide detailed information regarding activities that may require remote file access. The scenarios have been reviewed by the scientists for accuracy. The scenario details are in the ECS User Scenario Notebook, Volume I (doc # 194-00311TPW).

2: The impact of remote file access on the network bandwidth requirements are dependent on two main criteria. The first dependency is on the access pattern from the client, and the second is the storage format of the data.
There are three basic distinct access patterns:
1) the client program accesses the entire file
2) the client program accesses a contiguous section of the file
3) the client program accesses fragmented parts of the file

If the client requests the entire file, there will be no network gain. If the client requests a contiguous section of the file, say half of it, the network traffic will be roughly half of the traffic required to transfer the entire file. If the client requests many fragmented parts of the file the resulting traffic will depend on the size of the fragments when compared to the block size of the file system.

If the client program attempts to access only various subsets of data, performance will depend on the raw data storage format. The access pattern will be fragmented access, and the size of the fragments will depend on the storage format. If the data is subsetted at the storage location then there will be a reduction in traffic directly related to the percentage cut.

ECS plans to have hooks in place to add the remote file access capability, if it is required and determined to be feasible.

3: The ECS contractor is currently conducting a trade study to prototype local DFS and NFS file sharing. This effort is aimed at characterizing and comparing the performance of DFS and NSF and to determine the vendor support/maturity of remote file access products. Preliminary results are expected to be available by end of July '95.

4. Same as item 3.

As indicated earlier, the use of remote file access can impact the inter-DAAC traffic and, hence, the bandwidth needed to support the data flows. However, the actual gains to be realized depend on vendor implementation of specific features such as local disk-based caching, replication, the nature of the file access/operation and data storage format. Results of the prototyping study can help get a better understanding of the impact and feasibility of using remote file access between DAACs.

Finally, as the inter-DAAC data transfer volumes for Release A CDR are small and the bandwidth required to support these flows is relatively modest resizing of these links for CDR is, in our opinion, not warranted. For the Release B traffic we are currently developing a technical white paper estimating the potential inter-DAAC data transfer volume reduction from subsetting. Remote file access would be one of the potential means of implementing subsetting.

Status **Closed** **Date Closed** **6/20/95** **Sponsor** **Herring**

****** **Attachment if any** ******